Supplementary Information of *CTransformer*

Materials and Detailed Methods

The Animal *C. elegans*, Cell Morphology Map, and Embryogenesis

The model organism *Caenorhabditis elegans*, *C. elegans*, is widely used to study cellular developmental biology during embryogenesis due to its highly transparent body and regular cellular development, including consistent patterns in cell lineages [1], timings of divisions [2], orientations of axes [3], and gene expression [4-6]. To track these cellular processes and patterns, previous studies used transgenic markers, like GFP or *mCherry*, to trace lineages and images of fluorescent nuclei to show cellular differentiation, migration, and interactions [7-12]. The fluorescent labels allow descendants of individual cells to be tracked, with the lineages showing the cell cycles of developmental patterns [2, 4, 13-15].

Linking cell morphology to cell identity during *in vivo* embryogenesis allows investigation of the regulation of cell-cell contacts, gene expression, and determination of cell fates during development [16-18]. Four-dimensional (4D), 3D + time (t), imaging is the basic technique that could achieve this by the construction of a series of time-lapse images of embryonic development showing cellular morphology and cell lineages [19-21].

Multicellular animals, metazoans, show a range of complex deformations of cellular morphology [15, 22-24], and the durations of cell cycles are highly controlled by the asymmetry in sizes of single-cells [2, 13, 25, 26], with cell junctions and adhesions playing important developmental roles [27-29]. Specification of cell fate [30], programmed migration [31], genetic regulation of cell shape, and cellular contacts [22, 32, 33], can all benefit from a morphological map. Tracing complete cell lineages needs images of fluorescently stained nuclei [12, 33-37]. While this signal could occupy a possibly redundant channel, it would bring extra phototoxicity. Accurate and effective tracing of the segmentation of cell shapes and lineages up to the 550-cell embryo (for *C. elegans*) from only fluorescently-labeled membranes would be desirable and could facilitate easier processing. Different contrast methods have been applied, but they do not allow manual editing in a reasonable time [12, 38], especially for late-stage embryos.

Topological Wasserstein Distance Loss

Transformer-based (attention module) deep neural network solves the local features learning problem. However, the existing loss functions in voxel-wise recognition (segmentation), like the binary cross entropy (BCE), mean square error (MSE), and soft Dice score loss in VNet [39], are not coherent with regularizing the cellular boundary and have no receptive field in global 3D images. To drive the network to not only geometric accuracy but also cellular shape level correct, we propose a topology-constraint loss integrating EDT, persistent homology (PH), and $p-$Wasserstein distance [40]. The proposed loss includes 2 parts: geometrical L_g and topological loss L_t:

$$L_\epsilon(Y_{pred}, Y_{gt}) = L_g(Y_{pred}, Y_{gt}) + w_\epsilon \times L_t(Y_{pred}, Y_{gt}) + \xi,$$

where ϵ represents the current training epochs, $Y_{pred}\ and\ Y_{gt} \in R^{(H\times W\times D)}$ are the predicted and ground truth cell membrane volumes respectively, and W_ϵ is the weight of topological loss according to the training epochs. ξ is the necessary constant term for all loss functions. The geometrical loss is calculated by MSE: $L_g(y_{pred}, y_{gt}) = \sum|\mu_{(h',w',d')} - \nu_{(h',w',d')}|$, where $\mu_{(h,w,d)}$ and $\nu_{(h,w,d)}$ are respectively the probabilities of every voxel in y_{pred} and y_{gt}. Topological information and features can be extracted and learned by the *TUNETr* when the prediction is close to the ground truth, so the geometrical loss is the main loss L_ϵ at the initial training part. w_ϵ is a consistent weight for topological loss depending on the complexity of the images (the number of topological structures).

L_t from the cubical persistent homology (PH) diagram D, is calculated to extract topological structural differences of the prediction and ground truth 3D images (Euclidean distance transformed). For the volumetric images, they would be transformed to cubical complexes for representing images in a set of vertices, edges, facets, and other higher-dimensional counterparts, which are basic topological elements. PH is counting the number of $p-$dimensional topology structures (β_k), also known as k^{th} *Betti numbers*

$D_{pred}^{(k)}$ and $D_{gt}^{(k)}$, including the birth and death time of β_k (ref the PH showing figures), in the probability map. In the 3D space, there are 3 −dimensional topological features, $\beta_0, \beta_1, \beta_2$ such as the number of connected components (1D), holes (2D), and hollow voids (3D, similar 3D structure in a torus). The PH diagram is defined as $D = \sum_{k=0}^{p} D^{(k)}$ of an image.

Wasserstein distance is $optimal\ transport$ for the $Monge - Kantorovich$ problem of assessing the path, "difference" and distance from persistent homology diagram D_{pred} of Y_{pred} to D_{gt} of Y_{gt} [41, 42]. L_t is derived from p −Wasserstein distance between probability maps ($p = 3$ for 3D images) with stability to noise [40, 42, 43]:

$$W_p\left(D_{pred}^{(k)}, D_{gt}^{(k)}\right) = \left(inf_{\eta: D_{gt}^{(k)} \to D_{pred}^{(k)}} \sum_{x \in D_{gt}^{(k)}}^{1} \left|\left|x - \eta(x)\right|\right|\right)^{1/p},$$

where $\eta(x)$ denotes a bijection from D_{gt} to D_{pred}. D_{gt} and D_{pred} have different β_k, the bijection denotes that the points of D_{pred} would match with the closest point in D_{gt}. If the extra points have no match in D_{gt}, these points would be compared with the diagonal. Therefore, for D_{pred} of the prediction Y_{pred} and D_{gt} of the manually annotated ground truth Y_{gt}, the topology-constraint loss is formulated as：

$$L_t = \sum_k^p W_p\left(D_{pred}^{(k)}, D_{gt}^{(k)}\right).$$

The geometrical loss term L_g penalizes globally voxel-wise mismatches between *TUNETr*'s predictions and the GT, and the topological term L_t encourages the predicted probability map to generate similar amounts and structures of topology persistent homology with the GT.

Semi-supervised Learning with Edge-enhancing Synthetic GT for Boundary-aware Model

Achieving machine-based predictions with accuracy similar to human-level annotations in 550-cell embryos remains a challenge. During embryogenesis, cells can undergo significant morphological changes, such as elongating and aligning in grid patterns, with deformations, particularly from spherical to elongated shapes, posing considerable challenges for standard DNN and cell instance generation algorithms. Skin cells form elongated shapes, move to the outer edge of the embryo, and cover other cells from the 350-cell stage. After the 500-cell stage, skin cells cover more than 60% of the embryo's surface and continue to grow and close. Apoptotic cells shrink and will be engulfed by a surrounding cell and embedded in other cells, giving highly dense membrane signals, which result in unusual and ambiguous areas that are difficult to generate. Tissues, like "Other" and "Neuron", also proliferate at small cell sizes. *CTransformer* is able to manage irregular shapes and nearly invisible regions.

To address these significant obstacles and failure of recognizing boundary membrane, insufficient annotated GT, weak generalization, and low accuracy for supervised learning in processing 3D + *t* fluorescence images, we integrate semi-supervised learning and utilize the large number of unannotated images in *TUNETr*. We trained *iTUNETr* from a few manual GT with nuclei prior knowledge, and *iTUNETr* synthesizes manually-annotated level synthetic GT from the massive remaining 3D images. From the multi-cell segmentation from *iTUNETr*, the clear binary cell membrane of time-lapse volumes was constructed with the shape-constraint module to select human-annotated level GT for *sTUNETr* (a synthetically semi-supervised learning model, a part of *TUNETr*). This module would filter the volumes whose predicted probability maps are far away from their real topological structures and triple the fluorescence signal intensity in the outermost membrane of the GT to combat serious signal weakening problems. *TUNETr* boosts the potential and performance of our model with only 71 labelled images. Even in the late-stage imaging domain, *TUNETr* provides human-level membrane recognition and cellular segmentation (Figures 2 to 4). As illustrated in the main text, for 16 evaluation GT, *CTransformer* achieves a 4.1% cell loss rate. And the improved parts are mainly located in the skin and apoptotic tissues. Moreover, considering some manually annotated errors, the Dice Score of 0.9 of *CTransformer* has reached human-level accuracy. This training strategy's nuanced handling of spatial information and its sensitivity to cellular boundaries ensure that the integrity of cellular morphology is meticulously preserved, thereby enhancing the overall accuracy and reliability of the segmentation process.

Transformer-based Large Model

TUNETr, a UNet-like Swin Transformer deep learning neural network, is inspired by previous work (Figure S1b), UNet, Transformer, Swin Transformer, and SwinUNETR [44-46], for 3D + t cellular membrane recognition. Contracting images to high-level

spatially relative feature maps, *TUNETr* consists of a stack of transformers as the encoder. Embedding 1D sequence input from a volumetric image, *TUNETr* partitions a 3D image $x \in R^{(H\times W\times D)}$ into $N = \frac{H\times W\times D}{h\times w\times d}$ one-dimensional (1D) sequences, $x_v \in R^{(h\times w\times d\times C)}$, also named patches or tokens, which represent a word in Transformer structure. Here, (H, W, D) and (h, w, d) denote the resolutions of the 3D image and each patch, and C is the number of embedding spaces of the feature map (feature size). Self-attention computation is conducted in non-overlapped $(h \times w \times d \times C)$ patches (tokens) and the small-sized patches for reasonable computational complexity, but it has no communication with other patches while losing global attention. To obtain significant global attention in density prediction, *TUNETr* applied the shifted window based self-attention in the volumetric semantic segmentation. One window (layer l) contains (M, M, M) patches evenly in N patches from a 3D image, and the next layer l+1, is shifted with $(\frac{M}{2}, \frac{M}{2}, \frac{M}{2})$ voxels. The previous and subsequent shifted windows, layer l and l+1, both containing $M \times M \times M$ patches, with window-based multi-head self-attention ($W\text{-}MSA$) and shifted $W\text{-}MSA$ ($SW\text{-}MSA$) [45, 47] is calculated by

$$\hat{z}^l = W\text{-}MSA\big(LN(z^{l-1})\big) + z^{l-1},$$
$$z^l = MLP\big(LN(\hat{z}^l)\big) + \hat{z}^l,$$
$$\hat{z}^{l+1} = SW\text{-}MSA\big(LN(z^l)\big) + z^l,$$
$$z^{l+1} = MLP\big(LN(\hat{z}^{l+1})\big) + \hat{z}^{l+1},$$

where the $\hat{z}^l$ and $\hat{z}^{l+1}$ denote the outputs of $W\text{-}MSA$ and $SW\text{-}MSA$ modules, providing cross-window (global attention of one volume) connections. Multilayer perceptron module (MLP) convolutes the data to a target resolution (number of channels) and linear normalization module (LN) avoids the vanishing and exploring gradient problems. *TUNETr* also utilizes cyclic and reverse cyclic shift in masking and padding the hierarchical window partitions [45], for efficient self-attention calculation without messing up the original image order. To keep the significant spatial inductive bias, self-attention is calculated by $Attention(Q, K, V) = Softmax\big(QK^T/\sqrt{b}\big)V$, where $Q, K, V \in R^{(h\times w\times d\times b)}$ are the query, key, and value metrics in the transformer, and b is the query/key resolution [45, 47].

The encoder of *TUNETr* sets (h, w, d) of each patch as (2 x 2 x 2), and a patch length is $2 \times 2 \times 2 \times 2 = 16$, with 2 nucleus and membrane channels of the 3D volumetric images. With raw input 3D images $x \in R^{(128\times 128\times 128)}$, the feature size of the embedding space C is 48, and the depths of the attention head is (2, 2, 2, 2). Swin Transformer Blocks (respectively 8 $W\text{-}MSA$ and $SW\text{-}MSA$ modules) [45], *TUNETr* encoder processes the input $x \in R^{(128\times 128\times 128\times 2)}$ as following stages (Swin transformer block): (64, 64, 64, 48), (32, 32, 32, 96), (16, 16, 16, 192), (8, 8, 8, 384), (4, 4, 4, 768). Here, keeping explainable hierarchical and consistent with UNet, a patch merging module reduces the spatial resolution by 2 while the channel number would be embedded and increased by 4, and a linear convolutional neural layer reduces the channel number by 2 for the output of each stage (Swin Transformer Block).

The corresponding decoder integrating residual layers, increasing the resolution of feature maps, and receiving skip connections from stages of encoder, the upper data flow from high level to low level semantic (understandable) feature map denotes as: (8, 8, 8, 384), (16, 16, 16, 192), (32, 32, 32, 96), (64, 64, 64, 48), (128, 128, 128, 48). The output feature map at the bottleneck is fed into a convolutional layer, and a Euclidean distance transform global regression module (EDT-GFR) for cellular membrane recognition (prediction of probabilities of voxels being cell membrane).

Generating Multi-nuclei Segmentation GT and Training the Model

We meticulously extracted embryonic nuclei tracing data from high-resolution 4D Fluorescence Imaging, as delineated in the Methods section in the main text. This process entailed the precise determination of nuclear centroids, identities, and volumetric parameters. For the annotation of the nuclei ground truth (GT) dataset, encompassing 1270 *in vivo* embryonic volumes, we employed a multi-nucleus segmentation protocol. Each nucleus was annotated in three-dimensional space using spherical representations, meticulously ensuring consistency in resolution (H×W×D) with the corresponding cell membrane images.

To address the challenge of nuclear overlap, particularly pronounced in datasets beyond the 150-time-point threshold, we innovatively adapted our methodology. This adaptation involved halving the radii of the nuclear spheres, a strategic modification that significantly enhanced the spatial resolution of individual nuclei within densely populated regions.

Further, for the processing of nuclei GT, we leveraged the star-convex polygon model, as conceptualized in StarDist [48], extending it to encompass 96 radial vectors. This approach facilitated the generation of a nuanced gradient map, capturing topological variances and the radial distances from each voxel to the nearest nuclear centroid. Integrating the advanced StarDist3D framework [49]

into our *TUNETr* system was pivotal. This integration was tailored for the nucleus-prompting mechanism, utilizing star-convex polygons with 96 radial vectors to render a topologically informed gradient map. This map was pivotal in conveying spatial relationships, delineating equidistant angular separations, and quantifying voxel-centric nuclear proximities.

To optimize training of the star-convex polygon model, we employed a hybrid loss function that synergistically combines Mean Absolute Error (MAE) and Binary Cross-Entropy (BCE). This dual-functional loss function was instrumental in refining voxel classification and probability prediction. To augment the fidelity of our model's output, we implemented an advanced Non-Maximum Suppression (NMS) algorithm. This algorithm was critical in curtailing redundancy within the predicted volumetric data, thereby ensuring the integrity and precision of our segmentation outcomes.

Running Dual-approach Strategy with Nucleus-prompting

Algorithm 1: Delaunay-clustering algorithm.

Require: I_{pred}, h_{thres}

get locally minima points P_{min} with h_{thres}

get whole-cell segmentation set $S(cell)$

$S(cell) = watershed(I_{pred}, P_{min})$

build graphical edges set $S(edges)$

$S(edges) = Delaunay(S(cell))$

while $edge$ in $S(edges)$ **do**

    sum all overlapped and connected components

    calculate set of edge weights ω_{edge}

end while

drop and merge the overlapped meshes

return regular cells in $S(cell)$

To conquer the challenging multicell object segmentation from the occasionally incorrect recognition of cell membrane signals, *TUNETr* utilizes the 4D multi-nuclei segmentation from StarDist3D and the Delaunay-clustering watershed algorithm (Figure S2 and Algorithm 1) in 4D membrane recognition (semantic segmentation) to generate robust and accurate multiple whole-cell regions (multi-cell segmentation).

Using independent location seeds, we utilized the Delaunay clean method from them, where pixel clusters are organized into graphs using Delaunay triangulation. Without binarizing the grayscale probability maps and finding explicit nuclei positions as local minima, which suffers from implicit membrane recognition results, this step of *TUNETr* groups points near cell margins and assigns weights to graph edges. Overlapping cells (crossed edges and vertices) are also filtered out to achieve a cleaner segmented map. Conclude with marker-watershed segmentation using the refined seeds.

TUNETr capitalizes on the advanced capabilities of the marker-seeded watershed algorithm [50, 51], to integrate experimentally-determined multi-nucleus segmentation into its computational framework. This integration is manifested in the refinement of the predicted probability map, hereafter referred to as Y_{pred}. The incorporation of multi-nucleus segmentation serves a dual purpose: firstly, it provides precise, automated localization and quantification of cellular dimensions, and secondly, it acts as a critical navigational aid for the segmentation of multiple cellular objects within complex morphological structures.

Method Implementation Details and Evaluation Experimental Setup

TUNETr is partly implemented with PyTorch [52], StarDist3D [49], SwinUNETR [46], and MONAI [53]. *m2nGAN* is partly built on 3D Cycle GAN [54, 55].

Both *TUNETr* and *m2nGAN* are trained on a Dell computation station with 2 NVIDIA RTX A6000 GPUs (48GiB), 2 Intel Xeon Gold 6248R CPUs (3.00GHz), and 1TB physical memory. GPU driver version is 470.199.02 and CUDA version is 11.4. Python is 3.9.16, with 1.10.1+cu111 torch, 0.10.1+cu111 torchvision, and 1.2.0rc7+22.g575954cb *MONAI*. The 3D *iTUNETr*, *sTUNETr*, and the StarDist3D model (all the trained DNN models) in the end-to-end method *TUNETr* (the proposed segmentation method), as well as other comparing SOTA DNN models, are trained with the same computational resources and augmentation operation. The membrane recognition models are trained with 5000 epochs, $160 \times 160 \times 160$ randomly cut volumes as inputs. By perturbing the pixels' intensity (uniform distribution, scaled with the half-open interval $[1,1.1)$ and shifted with $[0,0.1)$) and randomly flipping (50% chance to be flipped along the x, y, and z axes, respectively), the 71 3D manually annotated volumes in the training dataset were augmented to multiple different and effective training 160×160×160 cube images. The Adam optimizer is used to update the network with an initial learning rate of 5×10^{-3} and a weight decay rate of 1×10^{-5}, using AMSGrad gradient descent optimization.

The evaluation experiments are conducted on 16 manually annotated volumes. The cell morphology of five embryos, "WT_C_Sample2", "WT_C_Sample3", "WT_C_Sample4", "WT_Sample1", and "WT_Sample7" was annotated by seven well-trained experts, from *Cshaper* [14], *CMap* [33], and the raw fluorescence images. They were re-grouped and organized. The segmentation results were manually checked and corrected slice by slice (2D view) and cell by cell (3D view), ensuring the correctness and smoothness of 3D cell shapes in the evaluation GT. Specifically, we gained new time-lapse and cell-number-wise GT data. First, the middle slice at each imaging time point throughout embryogenesis (255 2D images for "WT_Sample1" and 205 for "WT_Sample7") was annotated for 2D comparison, providing the cross-section of 30509 cell regions in total. Second, the complete typical 3D volume within 100±5−, 200±5−, 300±5−, 400±5−, 500±5−, and 550±5−cell stages (2 late-stage 3D images for either "WT_Sample1" or "WT_Sample7", 3 early stage for "WT_C_Sample2", "WT_C_Sample3", or "WT_C_Sample4") were annotated for 3D comparison, providing the full morphology of 4046 cell regions in total.

Additional Review and Results

Deep Learning-based Related Works

Manual annotation of 3D volumetric images is labor-intensive and time-consuming, making obtaining sufficient large-scale time-lapse cell morphological data for training impractical [14, 33, 35, 45], and a suitable training dataset is not available. Late-stage developmental embryos contain many cells with deformed shapes and anisotropic sizes, further complicating building a single-cell morphology map. At late stages, a slight recognition error of a cell membrane might cause a large area of cells to be incorrectly connected or lost, with possible exponential amplification of the error. In low-resolution cellular images, existing DNN methods, the data-driven models, could be improved to recognize the membrane signals in data with a different distribution.

However, as cells divide, nuclei are positioned more closely together in cells with small sizes, which can give unbalanced, ambiguous signals. Frequent and prolonged imaging is necessary to trace cells, which leads to photobleaching, the loss of fluorescent intensity over time, and phototoxicity that might damage cells and disrupt development. In *C. elegans*, the expression of a fused histone and marker protein can be too weak to detect in sublineages, like the progenies of D and E cells. Stronger expression might kill the embryo, but reducing the laser intensity gives no sufficient signals. Thus, the balance between signal quality and embryonic health influences the ability to establish and quantitatively analyze a map of the patterns of cellular morphology and gene expression.

Through balancing tuned images, advanced optical and deep learning segmentation algorithms have been established [14, 33, 35, 56, 57], however, their ability to recognize cell membrane signals and reconstruct morphological maps still has limitations. Under normal circumstances, the maps are derived from images of fluorescently labeled membranes and nuclei that give cell positions and detailed cellular shapes. However, occasional laser attenuation during imaging, insufficient 3D manually annotated ground truth (GT) data to develop a deep learning model, and the inability of a deep neural network (DNN) to recognize weak local signals remain problems.

Fully supervised convolutional and transformer-based DNNs [39, 44, 46, 47, 58-61], were proposed for pixel or voxel recognition (semantic segmentation), while the literature has not extensively tackled the above-mentioned recognition and segmentation

challenges. The first 5 DNNs, U−Net, VNet, 3DUNet++, DMFNet, and nnU-Net, are based on a convolutional neural network (CNN) structure, which encodes feature maps with spatially multi-scale communication and locality inductive bias, requesting a few ground truth (GT) for biomedical imaging analysis. The U-shaped DNNs expanded the depth of the neural network. The latter 3 DNNs, UNETR, SwinUNETR, and SwinUNETR−V2, are based on transformer and swin-transformer, create a local content-aware attention mechanism to leverage the context of the input patches (words or tokens in natural language processing, NLP). The breadth of the neural network has been enlarged by the transformer. Then, for multiple cell objects segmentation, competitive methods were developed [35, 57, 62-64], involving training one or a collection of specialized deep neural networks (DNNs), each tailored to a distinct imaging condition or animal as a "specialist" approach. "Specialist" DNN could not be extended well and directly to other imaging situations. Some "generalist" DNN approaches are proposed for general single-cell segmentation [48, 49, 65-67], to obtain multi-cell segmentation. They employ semantic segmentation of the cell membrane and instance segmentation as post-processing. StarDist and Cellpose are designed and adapted for 2D fluorescence microscopies [48, 65]. StarDist3D and Cellpose3D are applied to 3D, transforming the image domain and generating gradient feature maps to create 3D cell instances [65, 66], and assisted by star-convex and polygon theories, detecting the voxel-wise outline of a cell and generating fixed instances using an interpretable algorithm [48, 49]. Unfortunately, these methods are still unable to recognize extremely fuzzy membranes and establish irregular cell morphology in low-resolution images of living embryos. These DNNs address the problem of fluorescence signal recognition in blurred and unrecognizable microscopic images. However, all of the above methods rely heavily on large amounts of annotated data; they can only play a role of identification and preliminary (semantic) segmentation for good-quality images, lacking the capability of pushing the method to noisy and ambiguous images. Moreover, they have not been fitted generally, or the models are too small to handle complicated situations.

Solving Intrinsic Fluorescence Heterogeneity and Axially Weak Boundary Signals Membrane Recognition

First, when we conduct fluorescence imaging and optical scanning of 3D embryos, we observe intrinsic light heterogeneity and weak, unbalanced signals. The fluorescence heterogeneity was shown in the first column of Figure S7. The brightness of cell membrane signals in the inner and middle layers was very different, because the inner cells are larger and therefore their signaling proteins (mCherry here) expressions were less densely dispersed. This intrinsic anisotropy would reduce a big inner cellular segmentation failure and severe chain cell loss. Second, for living embryos, the laser beam was always tuned down for ordinary embryonic development, which would cause weak and ambiguous signals somewhere. The deficiencies in these images can be particularly severe in some areas (Figure S7). Meanwhile, restricted by the 3D scanning confocal imaging (2D slice by slice), the axial direction (perpendicular to the shooting plane) was in low resolution. It was difficult to maintain the spatial consistency of a living embryonic volume. Additionally, the areas circled by the large white rectangular box were the areas where the cell signal was severely blurred or disappeared, mainly in the last few layers of the slice for pseudo-3D imaging (Figure S7).

We proposed the topological loss function, cooperating with the large-scale model and semi-supervised and edge-enhancing learning for the cellular segmentation of 4D live-cell embryos, to solve these imbalances by producing membrane recognition with consistent recognized results (Figure 3a). Compared with images of other organisms, the membrane recognitions were in much lower anisotropic brightness distribution, no matter axially or sagittally. The signal strengths recognized by the inner, middle, and outer cell membranes of the embryo are relatively balanced. Moreover, the membrane recognition of *CTransformer* (Figure 3a), presents its comprehensive 3D membrane recognition for laser attenuated embryonic volumes.

For the first point, low-intensity and contrast signals also exist in the outermost layer of the cell membrane. This is due to two reasons. The first is because the inner cell membranes are in contact with each other, and there are two cell membranes with fluorescent protein expression; whereas the outermost cell membrane is in contact with the outside environment, which is only one layer of cell membrane, and the protein expression is only half of that on the inside, nearly disappeared (1st column of Figure 3b). This makes it very confusing for the DNN model to know if there are membranes. The second reason is that when the optical equipment scans the image layer by layer axially, there is a serious light scattering problem in the start and end layers of the scanning (Figure 3a). This is also a problem that image restoration and super-resolution in post-processing are trying to solve, but they have not conquered it (for

general images). *TUNETr* is effective in solving this weak boundary signal problem, as the outermost cell membrane is difficult to recognize. Thanks to the transformer-based and large-scale model, our model can pay closer attention to thickening the membrane contacting with the environment, as shown in the 6th column of Figure 3. Additionally, with the above-mentioned topological loss, *iTUNETr* connects the incomplete and invisible membrane as well, by learning the cell membrane skeleton of the whole embryo in the topological latent space, which is relatively intact, rather than fragmented as in other models.

Furthermore, for the second point, we are presenting the capability of *sTUNETr* for generating invisible membranes and axially undersampling signals. Qualitatively, *sTUNETr* allows extract foreground (membrane), without uncertain prediction around the membrane skeleton. For the axially (z direction) low-resolution issue (Figure 3a), semi-supervised learning produced a lot of pseudo-GT data with topological priors, to provide detailed 3D cellular morphological information axially. With a few annotated images and sufficient pseudo-training data, *sTUNETr* reconstructed the 3D cell membranes in ambiguous images. At late stage time points, cells become extremely small, containing only less than 10 voxels or 3 axial pixels, in some cases. Thus, *sTUNETr* improves axially undersampling signals, significantly enhances the constructions of not only ordinary cells but also irregular cell shapes, which are everywhere in large numbers in late-stage embryos. For laser-attenuated areas, they can be successfully recognized and segmented (Figure S7).

References

1. Sulston, J.E., et al., *The embryonic cell lineage of the nematode Caenorhabditis elegans.* Developmental biology, 1983. **100**(1): p. 64-119.
2. Fickentscher, R., P. Struntz, and M. Weiss, *Setting the Clock for Fail-Safe Early Embryogenesis.* Physical Review Letters, 2016. **117**(18).
3. Sugioka, K. and B. Bowerman, *Combinatorial Contact Cues Specify Cell Division Orientation by Directing Cortical Myosin Flows.* Developmental Cell, 2018. **46**(3): p. 257-270.e5.
4. Ma, X., et al., *A 4D single-cell protein atlas of transcription factors delineates spatiotemporal patterning during embryogenesis.* Nature Methods 2021 18:8, 2021. **18**(8).
5. Boyle, T.J., et al., *AceTree: a tool for visual analysis of Caenorhabditis elegans embryogenesis.* BMC Bioinformatics 2006 7:1, 2006. **7**(1).
6. Murray, J.I., et al., *Automated analysis of embryonic gene expression with cellular resolution in C. elegans.* Nature Methods, 2008. **5**(8): p. 703-709.
7. Bao, Z., et al., *Automated cell lineage tracing in Caenorhabditis elegans.* Proceedings of the National Academy of Sciences, 2006. **103**(8): p. 2707-2712.
8. Kretzschmar, K. and F.M. Watt, *Lineage Tracing.* Cell, 2012. **148**(1-2): p. 33-45.
9. Mcdole, K., et al., *In Toto Imaging and Reconstruction of Post-Implantation Mouse Development at the Single-Cell Level.* Cell, 2018. **175**(3): p. 859-876.e33.
10. Malin-Mayor, C., et al., *Automated reconstruction of whole-embryo cell lineages by learning from sparse annotations.* Nature Biotechnology 2022 41:1, 2022. **41**(1).
11. Fire, A., *A four-dimensional digital image archiving system for cell lineage tracing and retrospective embryology.* Bioinformatics, 1994. **10**(4): p. 443-447.
12. Sugawara, K., Ç. Çevrim, and M. Averof, *Tracking cell lineages in 3D by incremental deep learning.* eLife, 2022. **11**.
13. Guan, G., et al., *Volume segregation programming in a nematode's early embryogenesis.* Physical Review E, 2021. **104**(5): p. 054409.
14. Cao, J., et al., *Establishment of a morphological atlas of the Caenorhabditis elegans embryo using deep-learning-based 4D segmentation.* Nature Communications, 2020. **11**(1): p. 6254.
15. Guan, G., et al., *Comparison between phase-field model and coarse-grained model for characterizing cell-resolved morphological and mechanical properties in a multicellular system.* Communications in Nonlinear Science and Numerical Simulation, 2023. **117**: p. 106966.
16. Rasmussen, J.P., et al., *Cell Interactions and Patterned Intercalations Shape and Link Epithelial Tubes in C. elegans.* PLoS Genetics, 2013. **9**(9): p. e1003772.
17. Goldstein, B. and J. Nance, *Caenorhabditis elegans Gastrulation: A Model for Understanding How Cells Polarize, Change Shape, and Journey Toward the Center of an Embryo.* Genetics, 2020. **214**(2): p. 265-277.
18. Pentinmikko, N., et al., *Cellular shape reinforces niche to stem cell signaling in the small intestine.* Science Advances, 2022. **8**(41).
19. Domcke, S. and J. Shendure, *A reference cell tree will serve science better than a reference cell atlas.* Cell, 2023. **186**(6): p. 1103-1114.
20. Qiu, C., et al., *A single-cell time-lapse of mouse prenatal development from gastrula to birth.* Nature 2024 626:8001, 2024. **626**(8001): p. 1084-1093.
21. Packer, J.S., et al., *A lineage-resolved molecular atlas of C. elegans embryogenesis at single-cell resolution.* Science, 2019. **365**(6459): p. eaax1971.
22. Chen, L., et al., *Establishment of Signaling Interactions with Cellular Resolution for Every Cell Cycle of Embryogenesis.* Genetics, 2018. **209**(1): p. 37-49.
23. Mittnenzweig, M., et al., *A single-embryo, single-cell time-resolved model for mouse gastrulation.* Cell, 2021. **184**(11): p. 2825-2842.e22.
24. Junyent, S., et al., *The first two blastomeres contribute unequally to the human embryo.* Cell, 2024. **187**(11): p. 2838-2854.e17.
25. Li, Z., Z. Zhao, and H. Yan, *A Method for Quantitative Analysis of Long-distance Cellular Communication in 3D Time-lapse in vivo Living Embryos*. 2023, Research Square Platform LLC.
26. Jankele, R., et al., *Physically asymmetric division of the C. elegans zygote ensures invariably successful embryogenesis.* eLife, 2021. **10**: p. e61714.
27. Maître, J.-L. and C.-P. Heisenberg, *Three Functions of Cadherins in Cell Adhesion.* Current Biology, 2013. **23**(14): p. R626-R633.
28. Lee, C.-Y., et al. *Automated cell junction tracking with modified active contours guided by SIFT flow*. in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. 2014. IEEE.
29. Asan, A., S.A. Raiders, and J.R. Priess, *Morphogenesis of the C. elegans Intestine Involves Axon Guidance Genes.* PLOS Genetics, 2016. **12**(4): p. e1006077.
30. Mello, C.C., B.W. Draper, and J.R. Priess, *The maternal genes apx-1 and glp-1 and establishment of dorsal-ventral polarity in the early C. elegans embryo.* Cell, 1994. **77**(1): p. 95-106.
31. Li, X., et al., *Systems properties and spatiotemporal regulation of cell position variability during embryogenesis.* Cell Reports, 2019. **26**(2): p. 313-321.e7.
32. Ho, V.W.S., et al., *Systems-level quantification of division timing reveals a common genetic architecture controlling asynchrony and fate asymmetry.* Molecular Systems Biology, 2015. **11**(6): p. 814.
33. Guan, G., et al., *Cell lineage-resolved embryonic morphological map reveals signaling associated with cell fate and size asymmetry.* Nature Communications, 2025. **16**(1).
34. Katzman, B., et al., *AceTree: a major update and case study in the long term maintenance of open-source scientific software.* BMC Bioinformatics, 2018. **19**(1).
35. Cao, J., et al., *CShaperApp: Segmenting and analyzing cellular morphologies of the developing Caenorhabditis elegans embryo.* Quantitative Biology, 2024.
36. Guan, G., et al., *EmbSAM: cell boundary localization and Segment Anything Model for fast images of developing embryos.* Communications Biology, 2025.

37. Huang, Z., Z. Li, and H. Yan, *A Fusion Model With Effective Multi-Scale Parallel Transformer for Cellular Segmentation.* IEEE Transactions on Computational Biology and Bioinformatics, 2025: p. 1-9.
38. Waliman, M., et al., *Automated cell lineage reconstruction using label-free 4D microscopy.* GENETICS, 2024. **228**(2).
39. Milletari, F., N. Navab, and S.-A. Ahmadi. *V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation*. in *2016 Fourth International Conference on 3D Vision (3DV)*. 2016. IEEE.
40. Panaretos, V.M. and Y. Zemel, *Statistical Aspects of Wasserstein Distances.* Annual Review of Statistics and Its Application, 2019. **6**(1): p. 405-431.
41. Clough, J.R., et al., *A Topological Loss Function for Deep-Learning Based Image Segmentation Using Persistent Homology.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022. **44**(12): p. 8766-8778.
42. Waibel, D.J.E., et al., *Capturing Shape Information with Multi-scale Topological Loss Terms for 3D Reconstruction*, in *Lecture Notes in Computer Science*. 2022, Springer Nature Switzerland. p. 150-159.
43. Cohen-Steiner, D., et al., *Lipschitz Functions Have L p -Stable Persistence.* Foundations of Computational Mathematics, 2010. **10**(2): p. 127-139.
44. Ronneberger, O., et al., *U-Net: Convolutional Networks for Biomedical Image Segmentation*, in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. 2015, Springer International Publishing. p. 234-241.
45. Liu, Z., et al. *Swin transformer: Hierarchical vision transformer using shifted windows*. in *Proceedings of the IEEE/CVF international conference on computer vision*. 2021.
46. Hatamizadeh, A., et al., *Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images*, in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. 2022, Springer International Publishing. p. 272-284.
47. Hatamizadeh, A., et al., *UNETR: Transformers for 3D medical image segmentation*, in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 2022. p. 574-584.
48. Schmidt, U., et al., *Cell Detection with Star-Convex Polygons*, in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. 2018, Springer International Publishing. p. 265-273.
49. Weigert, M., et al. *Star-convex polyhedra for 3D object detection and segmentation in microscopy*. in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 2020.
50. Li, D., et al., *An Edge Embedded Marker-Based Watershed Algorithm for High Spatial Resolution Remote Sensing Image Segmentation.* IEEE Transactions on Image Processing, 2010. **19**(10): p. 2781-2787.
51. Van Der Walt, S., et al., *scikit-image: image processing in Python.* PeerJ, 2014. **2**: p. e453.
52. *Paszke, A.*, et al., *PyTorch: An Imperative Style, High-Performance Deep Learning Library*, in *Advances in Neural Information Processing Systems (NeurIPS 2019)*. 2019.
53. Cardoso, M.J., et al., *Monai: An open-source framework for deep learning in healthcare.* arXiv preprint arXiv:2211.02701, 2022.
54. Ge, Y., et al. *Unpaired Mr to CT Synthesis with Explicit Structural Constrained Adversarial Learning*. in *IEEE 16th International Symposium on Biomedical Imaging (ISBI)*. 2019. IEEE.
55. Isola, P., et al. *Image-to-image translation with conditional adversarial networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
56. Li, Z., et al., *An effective method for quantification, visualization, and analysis of 3D cell shape during early embryogenesis.* Quantitative Biology, 2025. **13**(1).
57. Cao, J., et al., *3DMMS: robust 3D Membrane Morphological Segmentation of C. elegans embryo.* BMC Bioinformatics, 2019. **20**(1).
58. Chen, C., et al., *3D Dilated Multi-fiber Network for Real-Time Brain Tumor Segmentation in MRI*, in *Lecture Notes in Computer Science*. 2019, Springer International Publishing. p. 184-192.
59. Zhou, Z., et al., *UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation.* IEEE Transactions on Medical Imaging, 2020. **39**(6): p. 1856-1867.
60. Isensee, F., et al., *nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation.* Nature Methods, 2021. **18**(2): p. 203-211.
61. He, Y., et al., *SwinUNETR-V2: Stronger Swin Transformers with Stagewise Convolutions for 3D Medical Image Segmentation*, in *Lecture Notes in Computer Science*. 2023, Springer Nature Switzerland. p. 416-426.
62. Stegmaier, J., et al., *Real-Time Three-Dimensional Cell Segmentation in Large-Scale Microscopy Data of Developing Embryos.* Developmental Cell, 2016. **36**(2): p. 225-240.
63. Azuma, Y. and S. Onami, *Biologically constrained optimization based cell membrane segmentation in C. elegans embryos.* BMC Bioinformatics, 2017. **18**(1): p. 307.
64. Azuma, Y., H. Okada, and S. Onami, *Systematic analysis of cell morphodynamics in C. elegans early embryogenesis.* Frontiers in Bioinformatics, 2023. **3**: p. 1082531.
65. Stringer, C., et al., *Cellpose: a generalist algorithm for cellular segmentation.* Nature Methods, 2021. **18**(1): p. 100-106.
66. Eschweiler, D., R.S. Smith, and J. Stegmaier. *Robust 3d Cell Segmentation: Extending The View Of Cellpose*. in *2022 IEEE International Conference on Image Processing (ICIP)*. 2022. IEEE.
67. Wang, A., et al., *A novel deep learning-based 3D cell segmentation framework for future image-based disease detection.* Scientific Reports, 2022. **12**(1).

Supplementary Figures

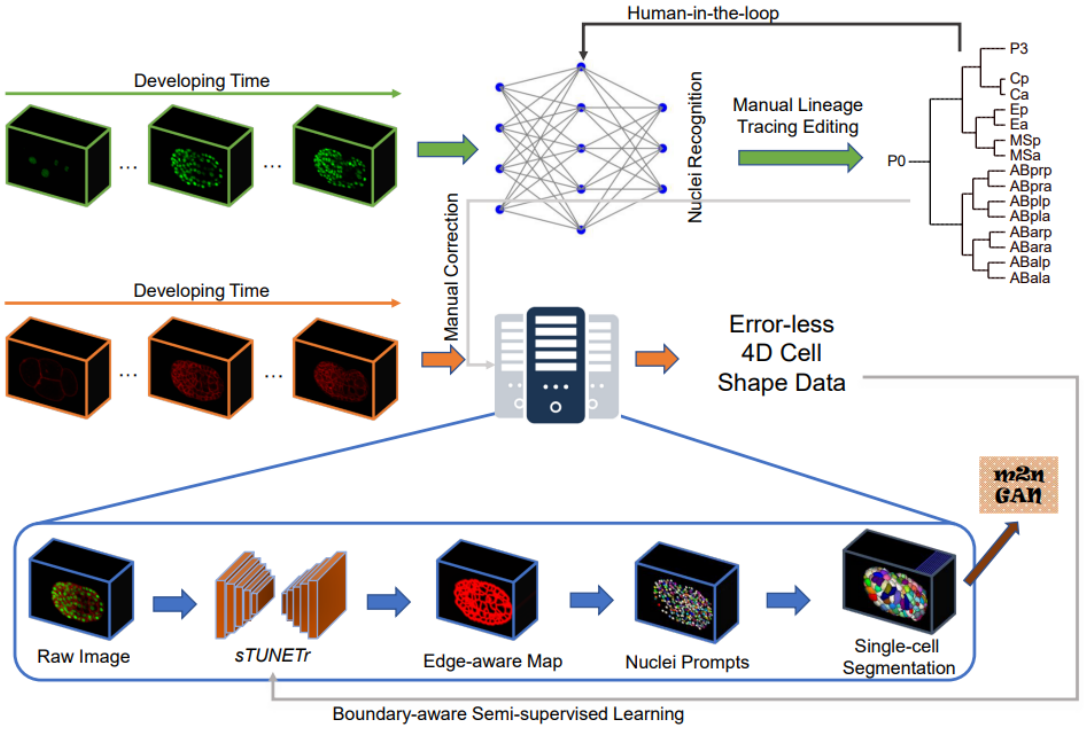


Figure S1. The framework of *CTransformer* for complete cell morphology map to 550−cell stage with nuclei images assisted. The human-in-the-loop workflow for semi-supervised learning strategy in training the single-cell segmentation process.

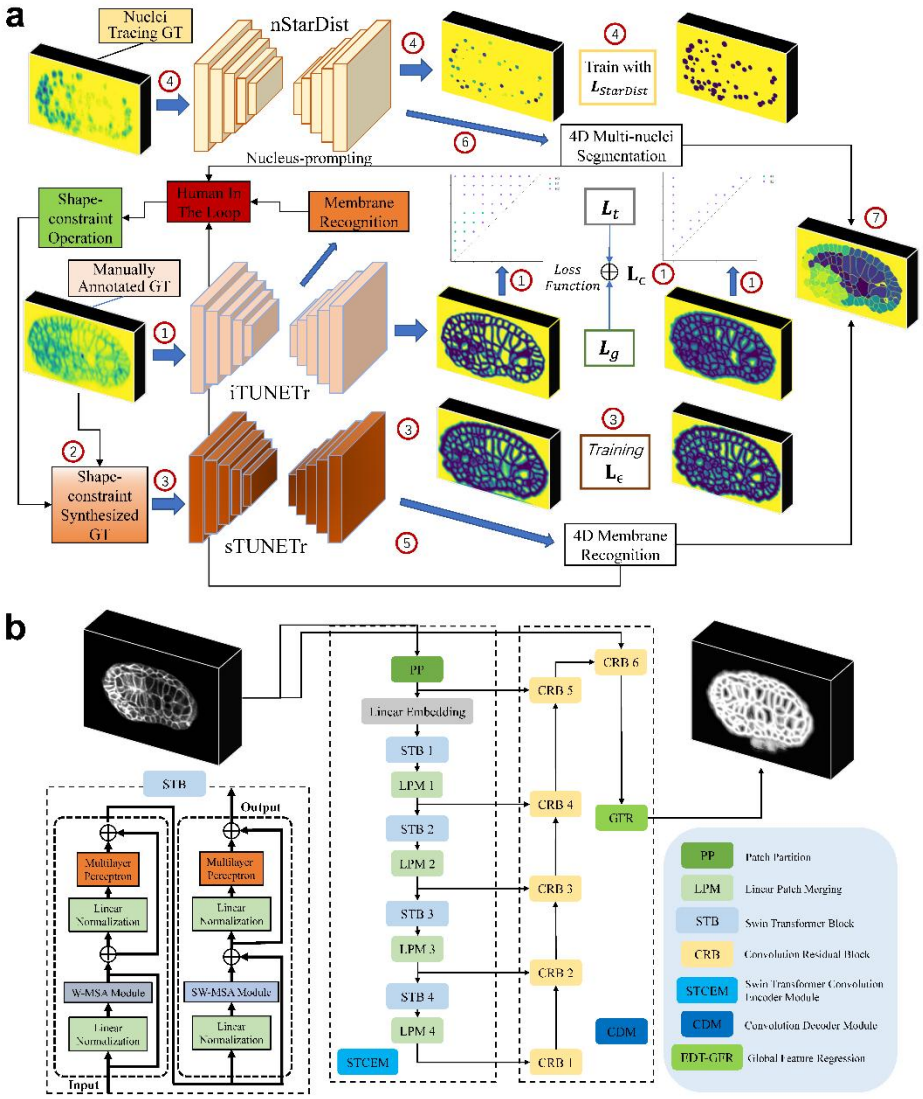


Figure S2. The training process for the integrated DNN model, *TUNETr* of *CTransformer* framework. (**a**) TUNETr of *CTransformer* is a well−designed DNN model for cell membrane recognition and multi−cell segmentation. In (**a**), from 1 to 3, and 5 steps, *CTransformer* trained 2 *TUNETr* models, *iTUNETr* and *sTUNETr* for 3D cell membrane recognition, and an adaptive StarDist3D model for 4D nuclei segmentation. In 4, 6 to 7 steps, *CTransformer* utilized the results of cellular membrane recognition, nuclei segmentation and Delaunay−watershed algorithm to produce the multiple whole−cell segmentation for 4D fluorescence images. (**b**) The figure illustrates the transformer, based on encoder−decoder DNN structure used in *iTUNETr* and *sTUNETr*. The most important part is the STB module, deploying attention mechanism in the DNN. The details of modules are explained in the Section Materials and Methods.

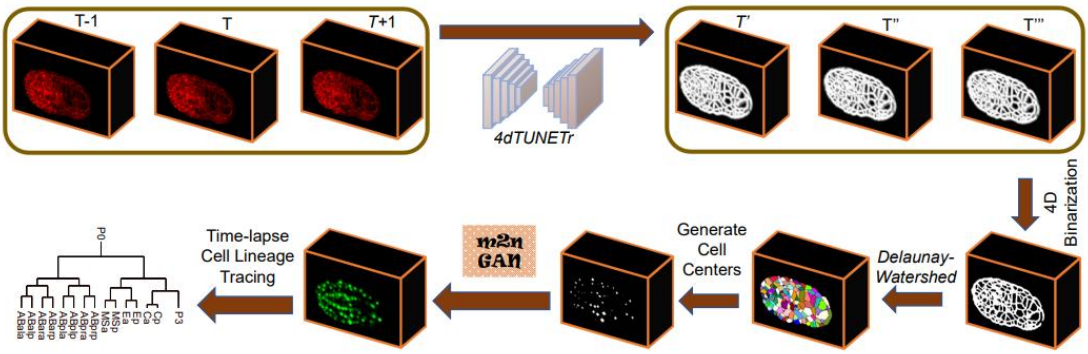


Figure S3. The *4dTUNETr* and *m2nGAN* detailed running procedure. *4dTUNETr* accepts 3 sequential raw membrane volumes as input and predicts the redundant membranes for one time point. Without cell shape centroids prompting, the *m2nGAN* generates pseudo nuclei images as the resource in lineage tracing.

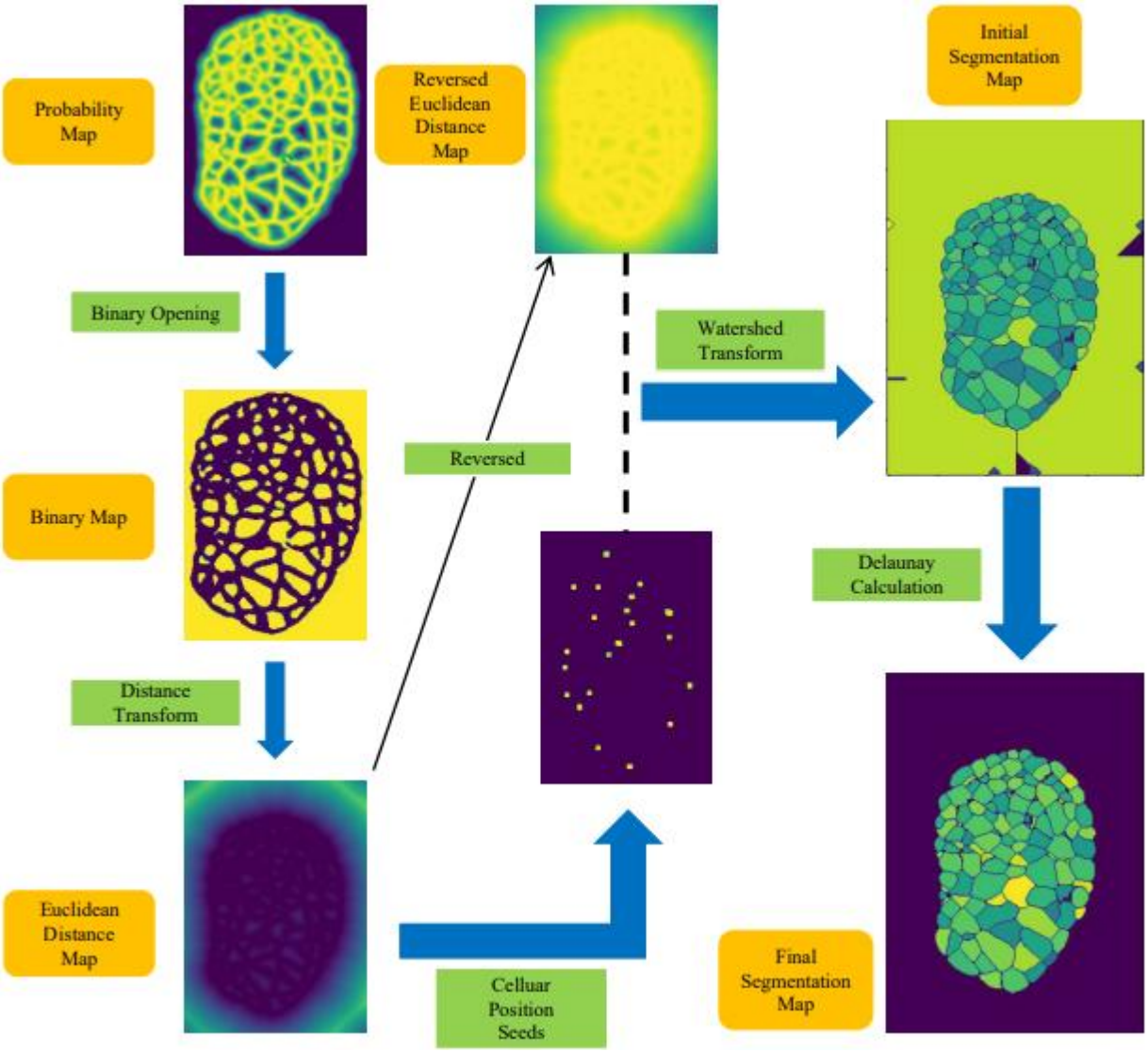


Figure S4: The complete 3D images processing dataflow of running output *TUNETr*, Delaunay−Watershed graphical algorithm without nuclei prompting. The dataflow is illustrated with 2D cut of the volumes. Cell membrane recognition (probability map prediction), binarized membrane segmentation, Euclidean distance transformed map, cellular position seeding map, multiple whole−cell segmentation, and Delaunay−processed segmentation are shown.

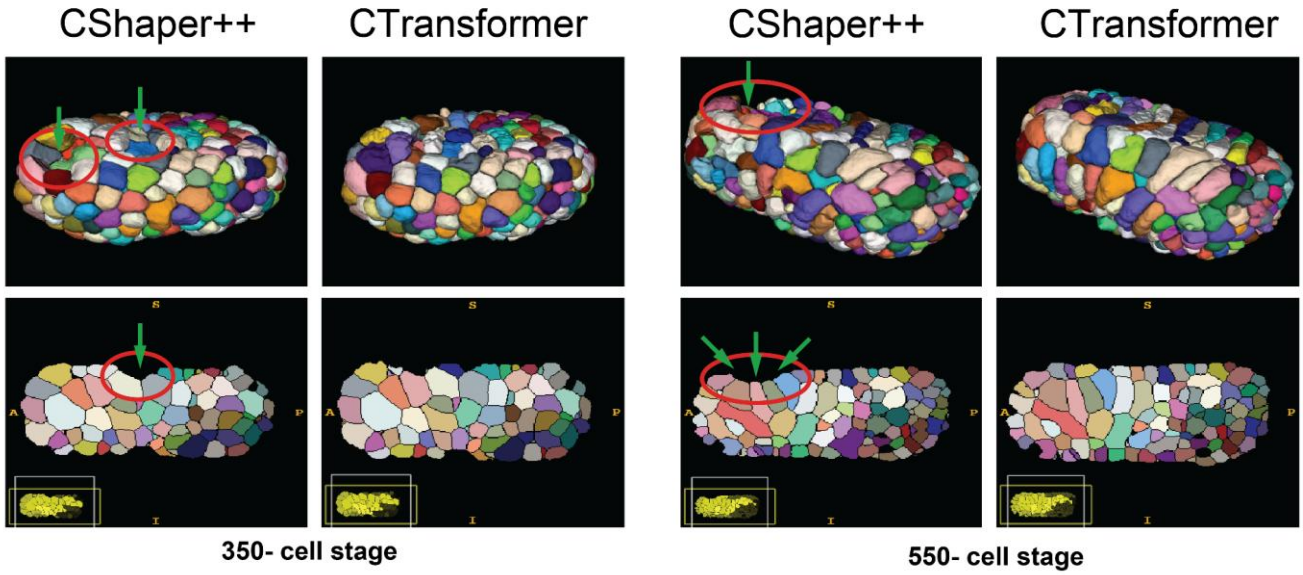


Figure S5: Qualitative comparison for the boundary area between *CShaper++* and *CTransformer*. Especially for the skin cells at the embryo back of the worm.

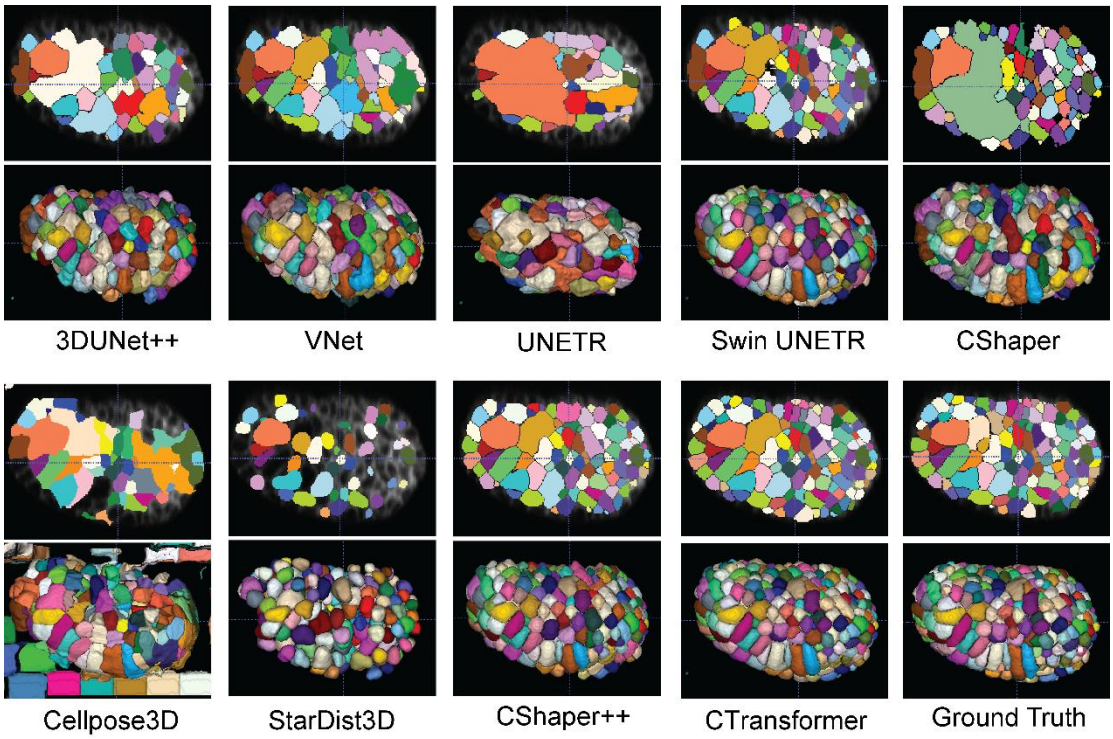


Figure S6: Qualitative illustration between *TUNETr* and other existing SOTA segmentation algorithms. 2D YZ view (upper row) and 3D side view (lower row) of the ground truth segmentation and segmentations generated from different DNN models, like VNet, UNETR, Cellpose3D, StarDist3D at 550−cell stage in *C. elegans*. Our results are generated by *sTUNETr*.

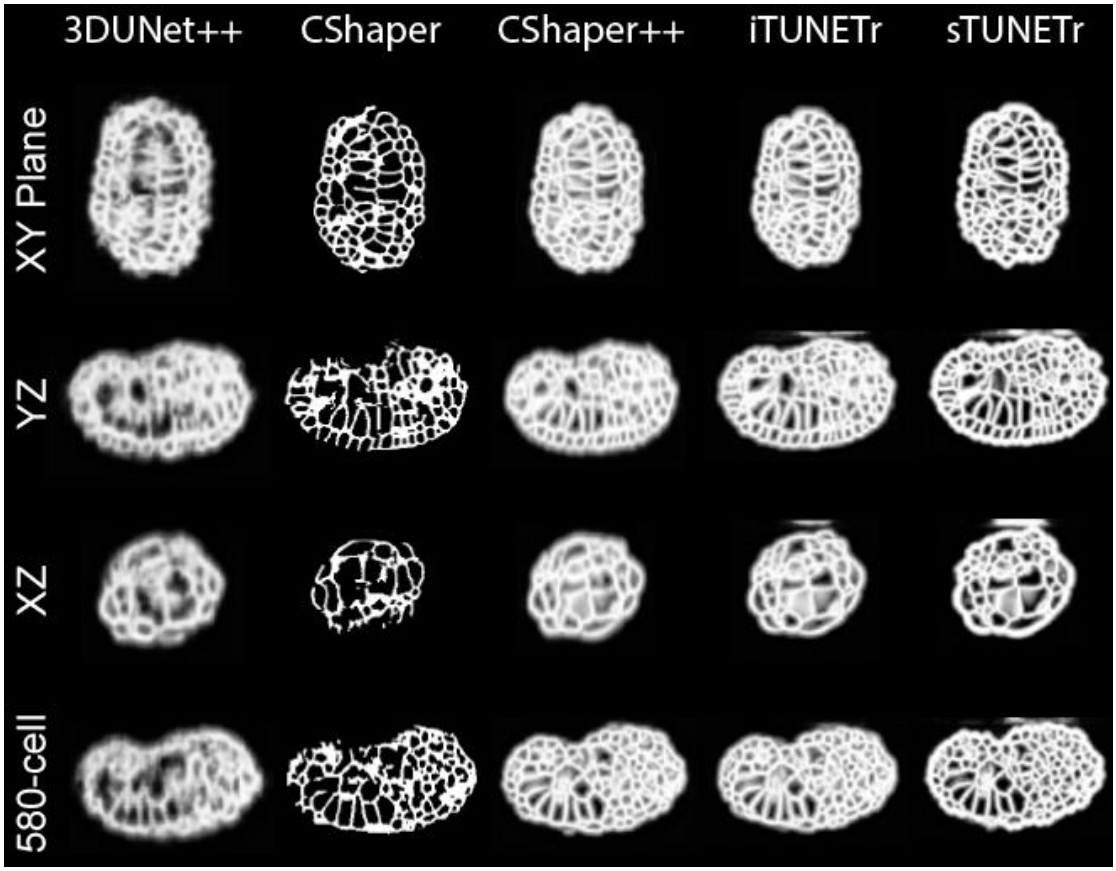


Figure S7. Qualitative demonstrations membrane recognitions at 550- and 580-cell stage. At XY (lateral) plane, first row shows the low membrane brightness because of intrinsic fluorescence heterogeneity. Axially, second row presents the significant loss of signal at the edge of the volume. Third row demonstrates the weak signals at bottom and weakened intensities along the laser direction. Fourth row, i.e. 580-cell stage embryo, illustrates severely fuzzy signals and undersampling from isotropic imaging. The five columns behind were giving membrane recognition results from 5 typical DNN models.

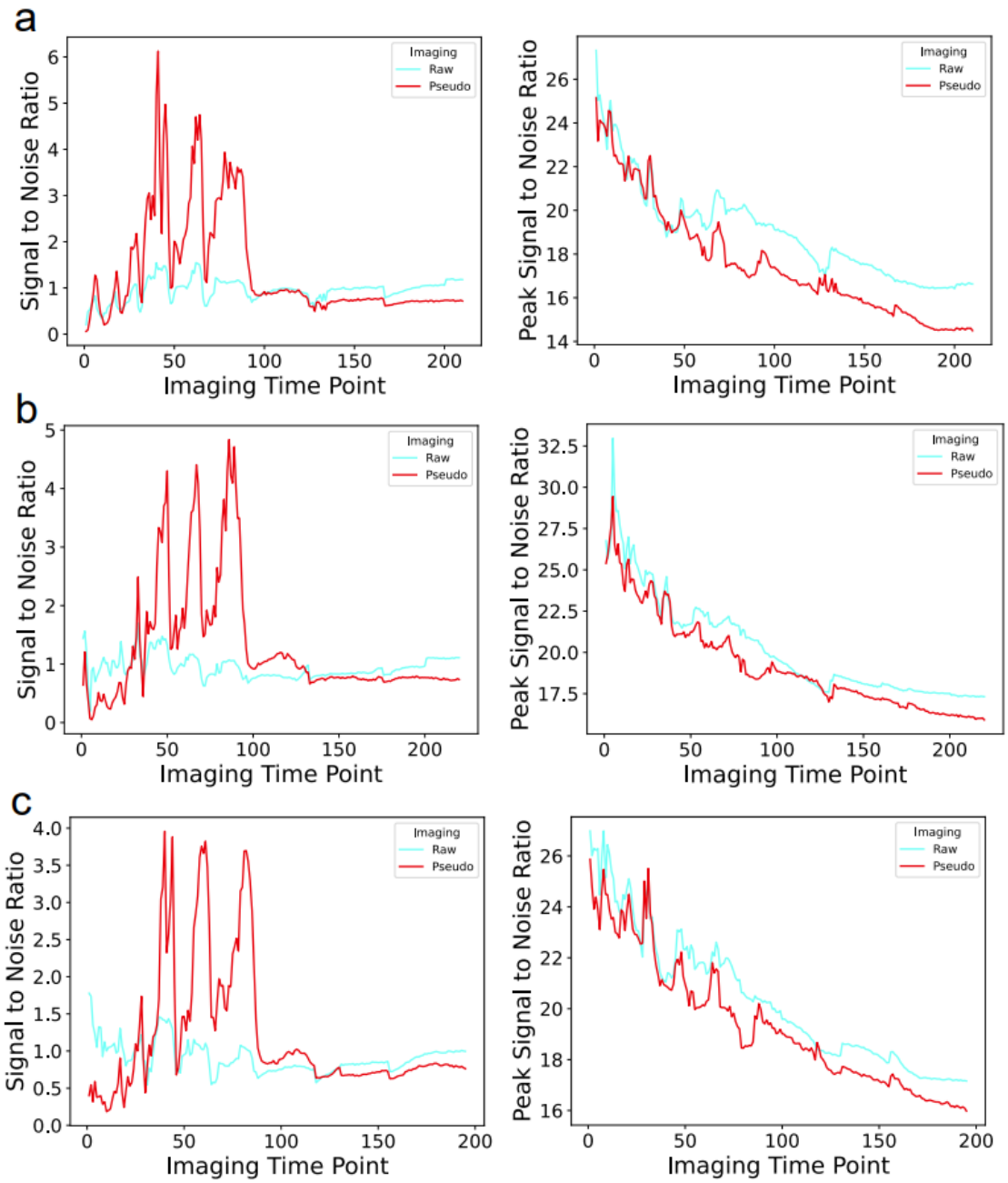


Figure S8: Image-wise evaluation of another two embryos between pseudo and raw nuclei images. Comparison of Signal to Noise Ratio (SNR) and Peak Signal to Noise (PSNR) between raw (cyan) and pseudo images (red) for one embryo across time. Shown are the values of the living embryo from 1 to 220 time point. Note that a higher SNR or PSNR depicts less signal discrepancies from ground truth images, and more signals and less noises for nuclei tracing. (**a**) The comparison of WT_C_Sample6. (**b**) The evaluation of WT_Sample4. (**c**) The results from WT_Sample5.

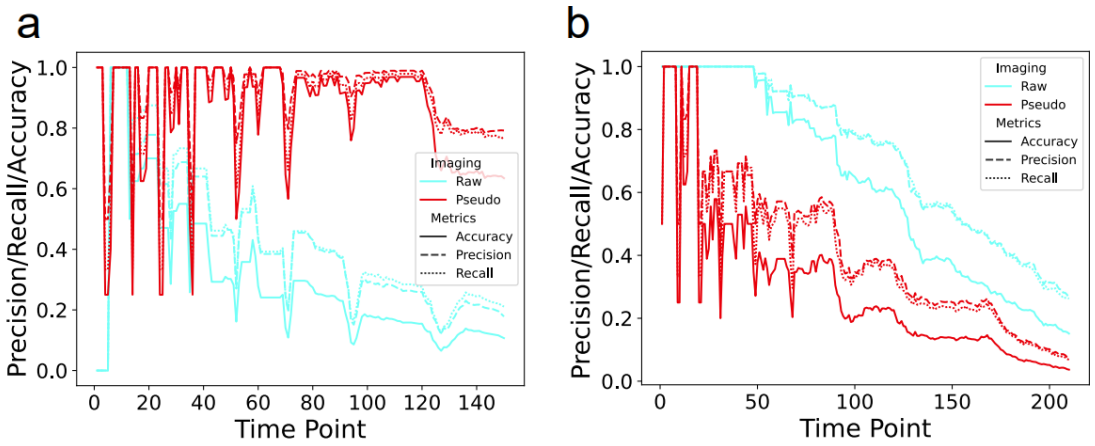


Figure S9: Lineage tracing comparison, accuracy, precision and recall between generative and raw nuclei images on embryos (a) WT_Sample4 and (b) WT_C_Sample6, respectively, with StarryNite.

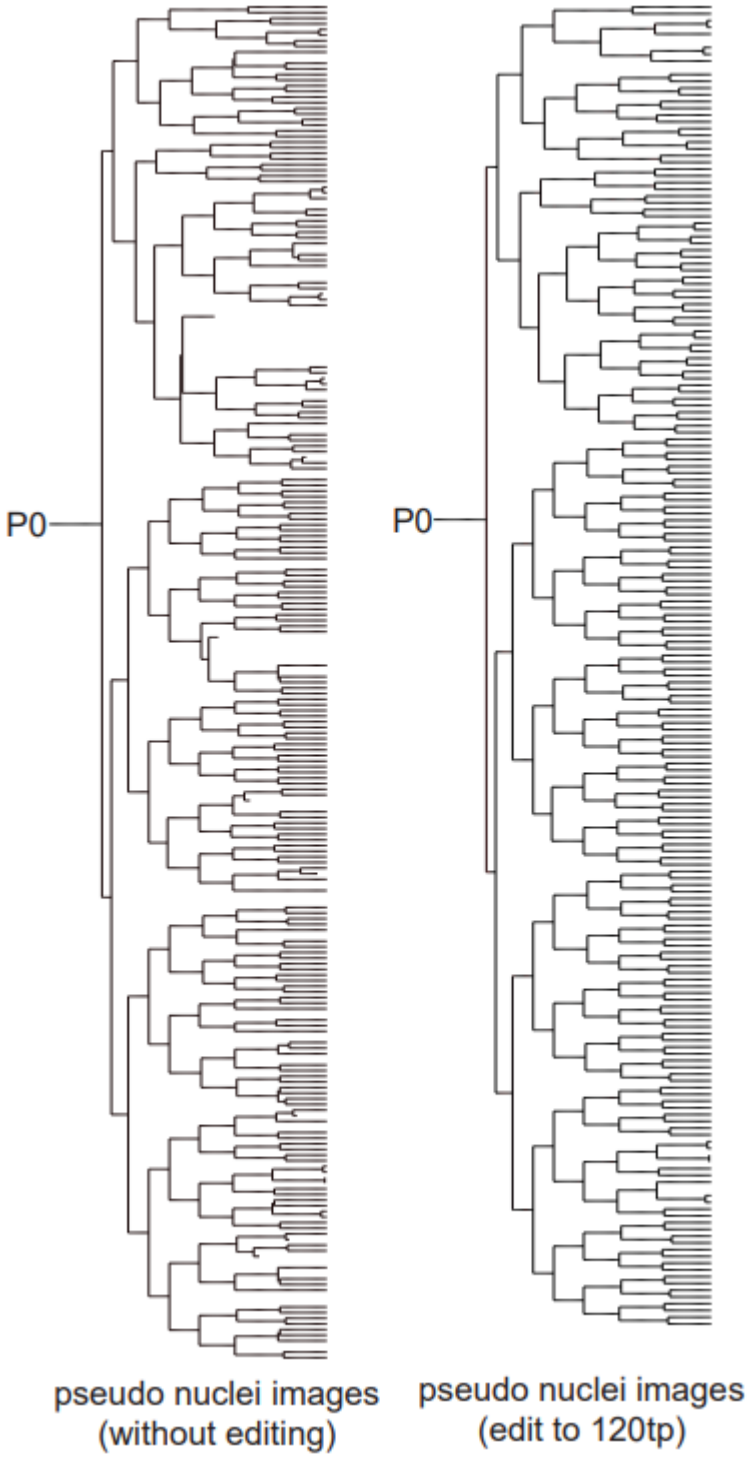


Figure S10: Before and after edited on generative pseudo nuclei images generated on WT_Sample4 with StarryNite and edited by AceTree. The edited cell lineage tree showcases the accuracy and facility of the produce linage tree with moderate number of tracking and tracing errors.

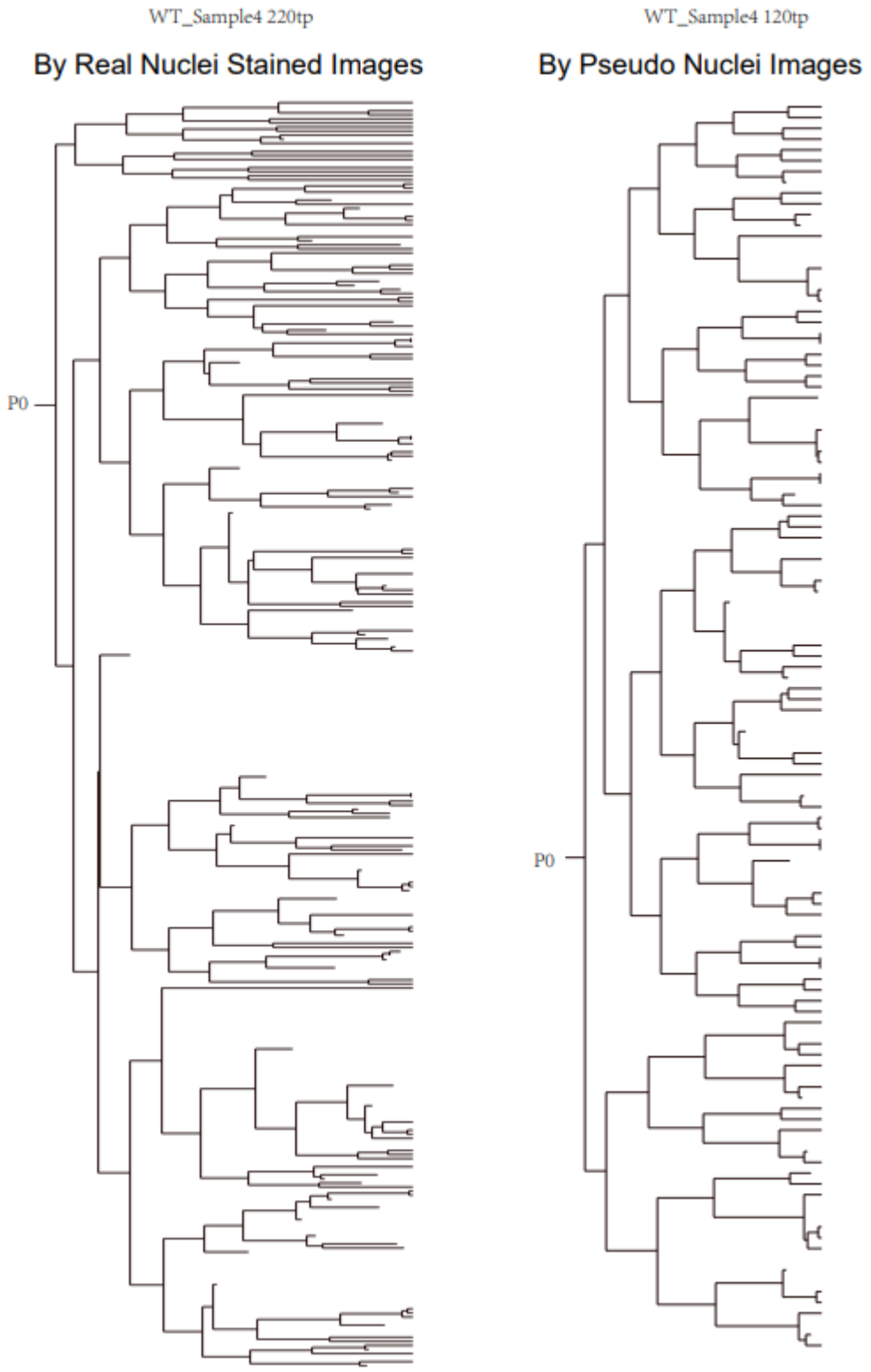


Figure S11: Lineage−wise (the tree) comparison between generative and raw nuclei images on uncompressed embryo WT_Sample4 with StarryNite and AceTree. P0 is the root and demonstrated to 120 time point without manual curation.

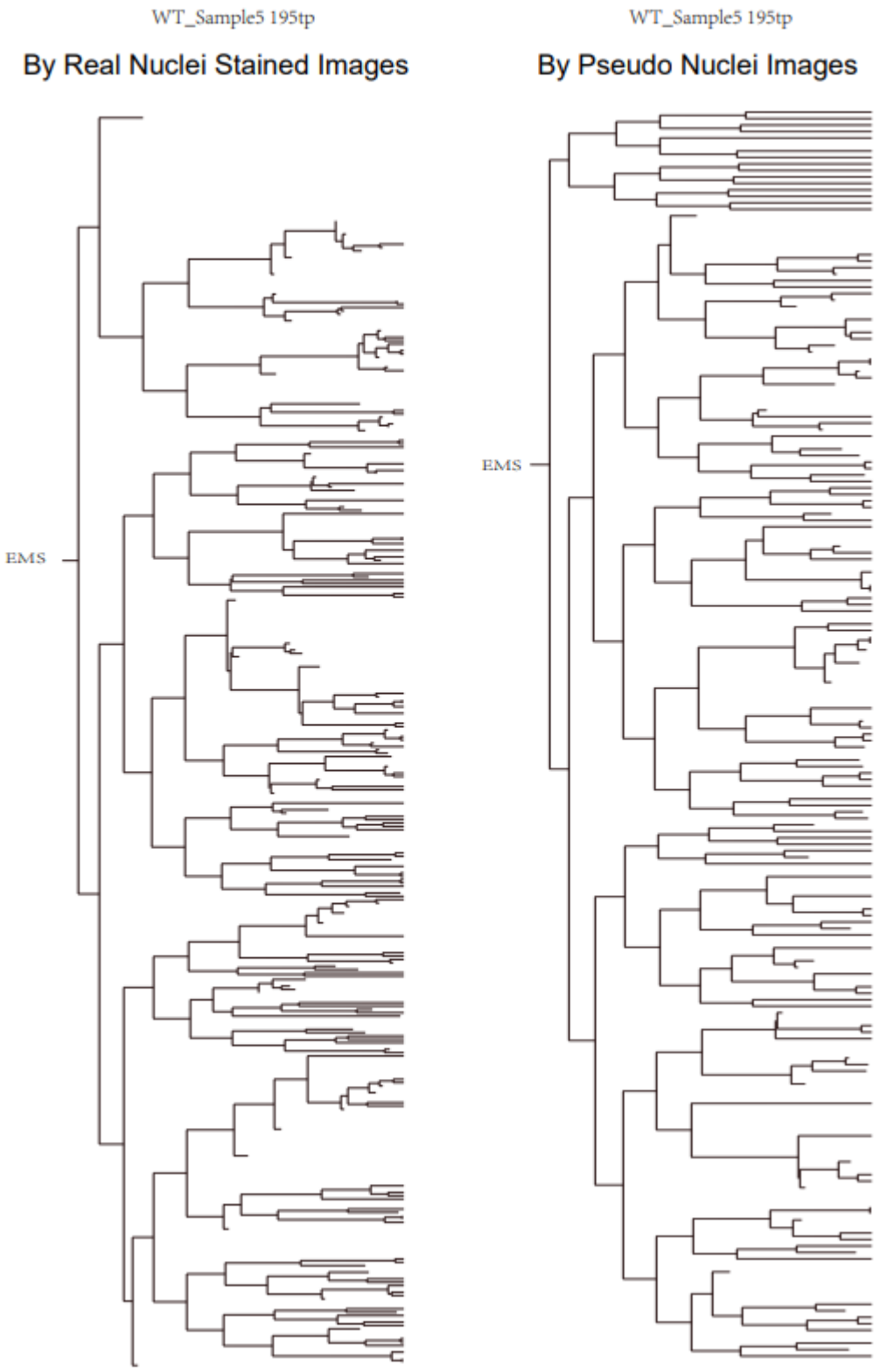


Figure S12: Lineage−wise (the tree) comparison between generative and raw nuclei images on uncompressed embryo WT_Sample5 with StarryNite and AceTree. EMS is the root and pushed to 195 time point, up to 550-cell stage without manual curation.

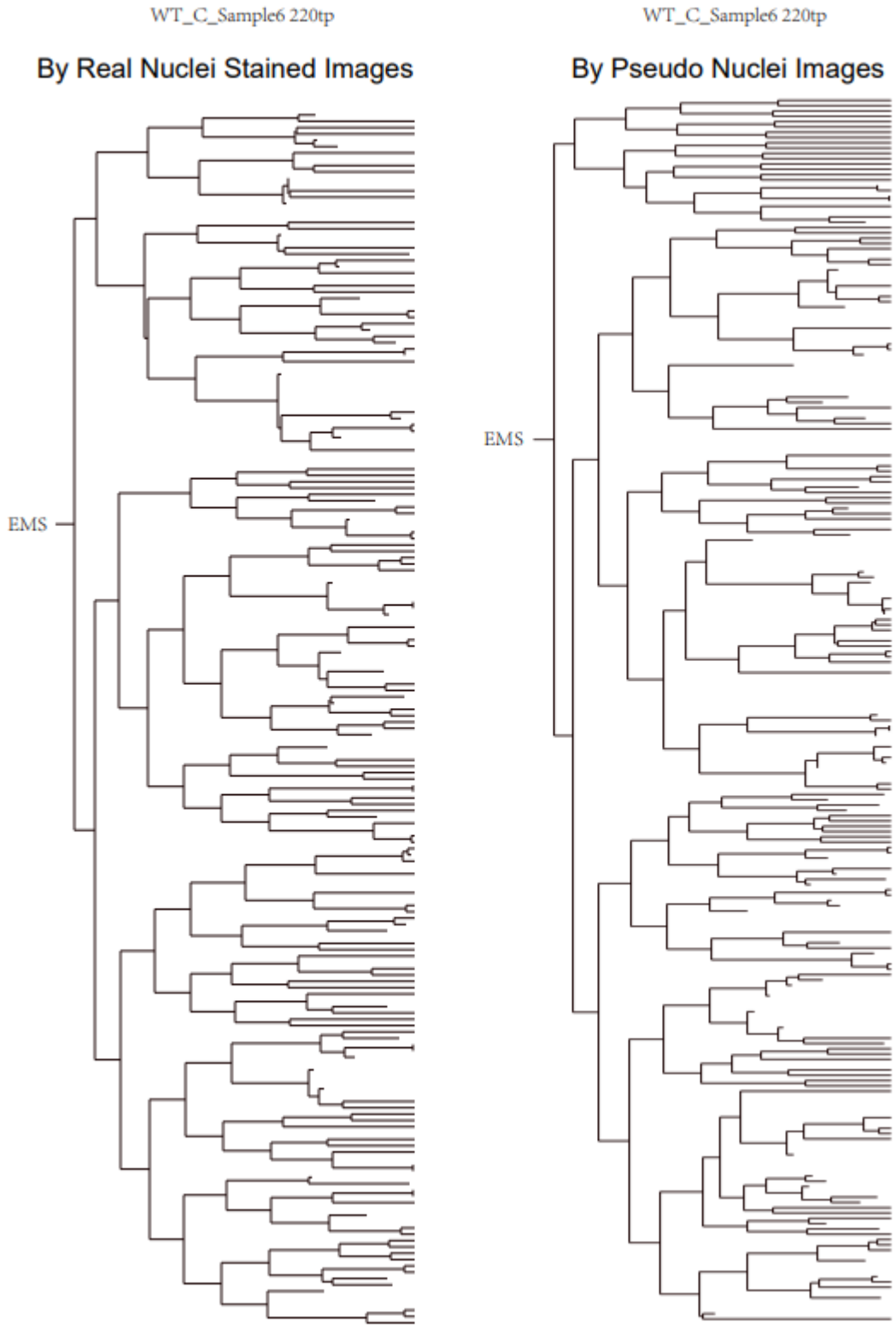


Figure S13: Lineage−wise (the tree) comparison between generative and raw nuclei images on compressed embryo WT_C_Sample6 with StarryNite and AceTree. EMS is the root and pushed to 210 time point, up to 530-cell stage without manual curation.

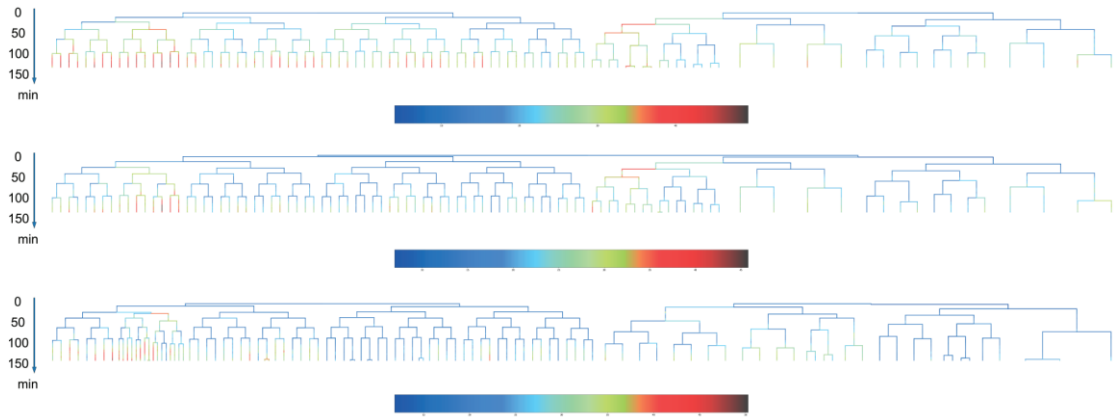


Figure S14. The cell-wise adhesion demonstrations of three time-lapse embryos (quantitative analyses on *hmr-1*).

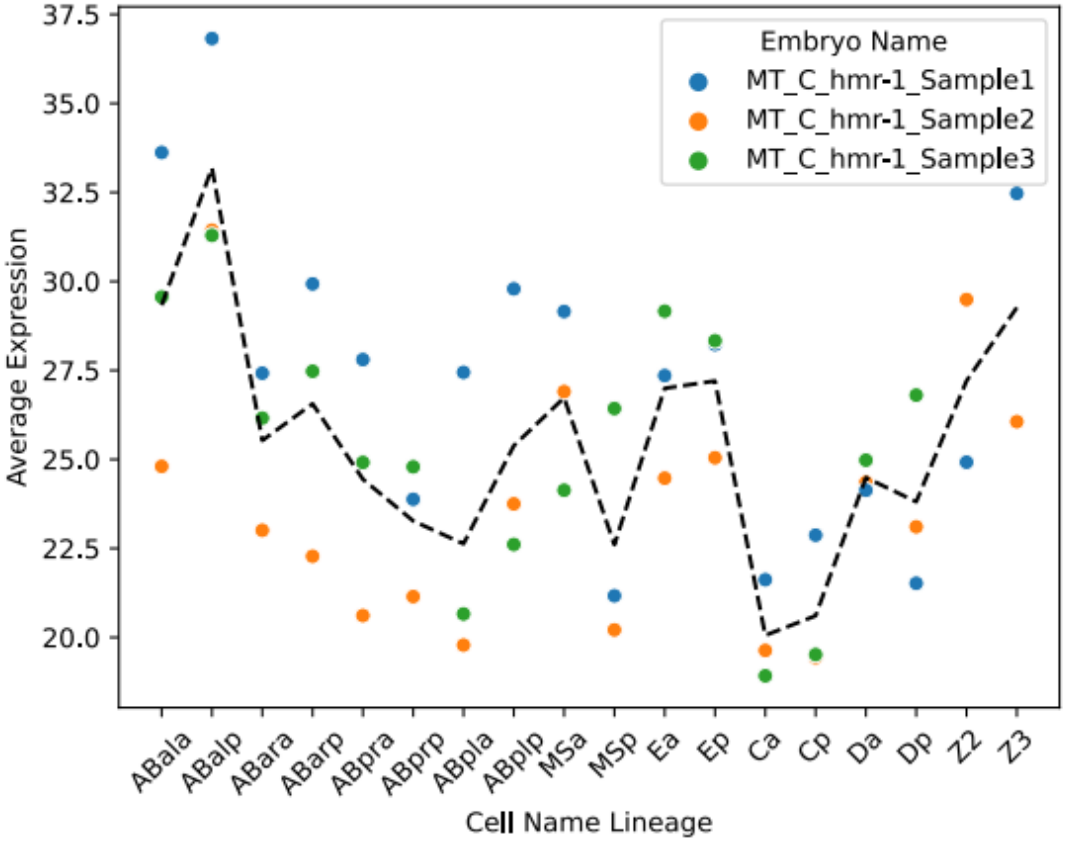


Figure S15. Surface unit cell adhesion on 18 sublineages. We could see the expressions among embryos are stable and regular.

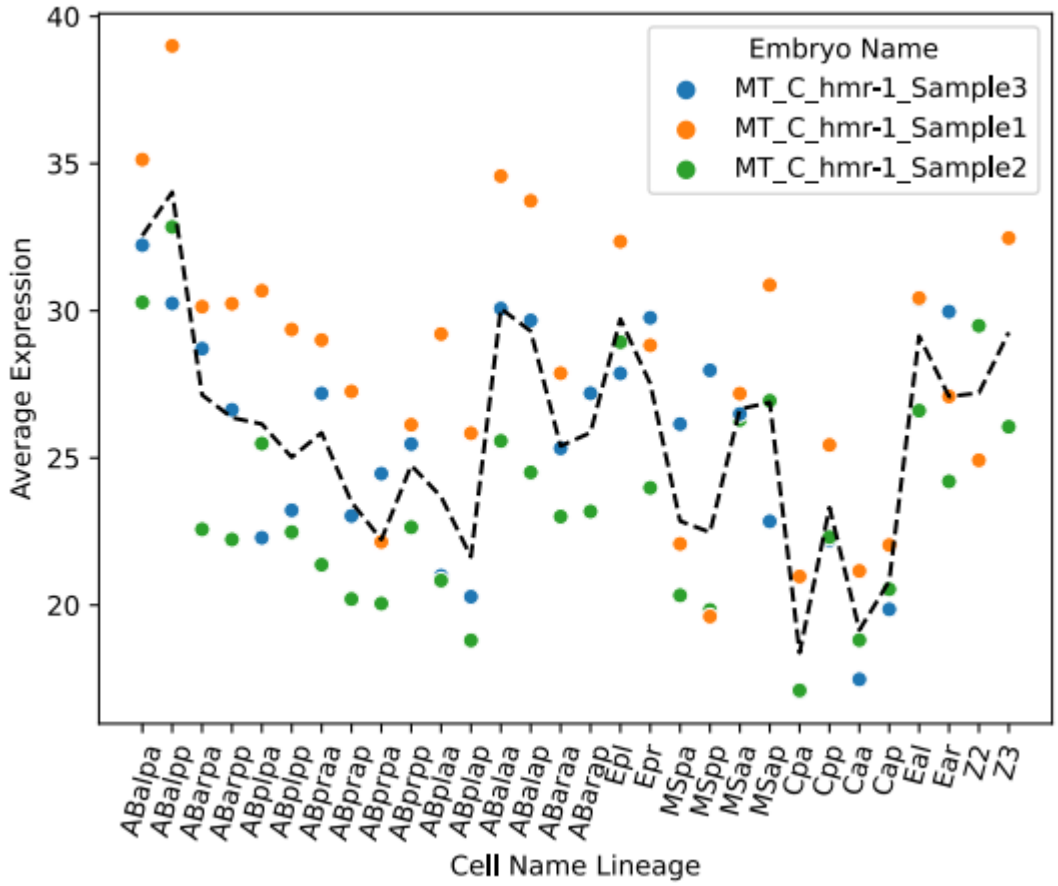
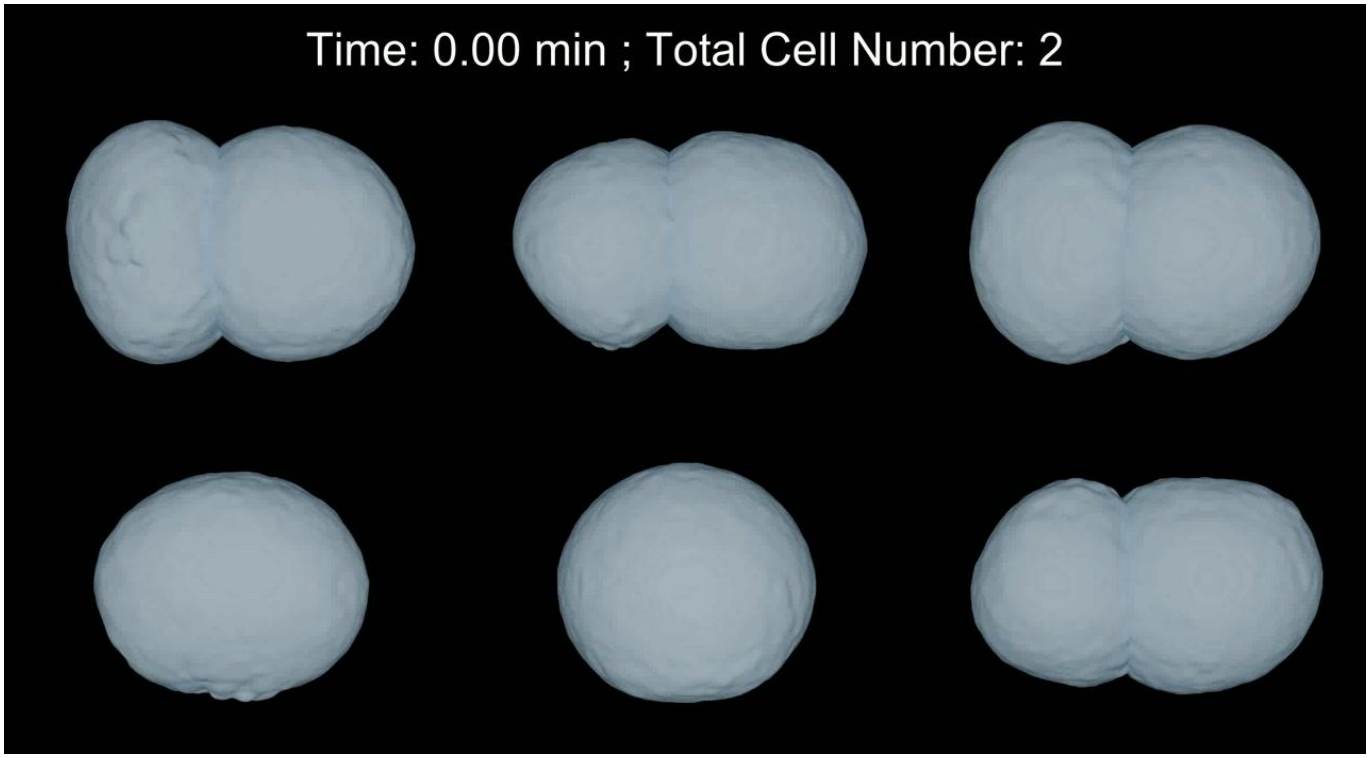


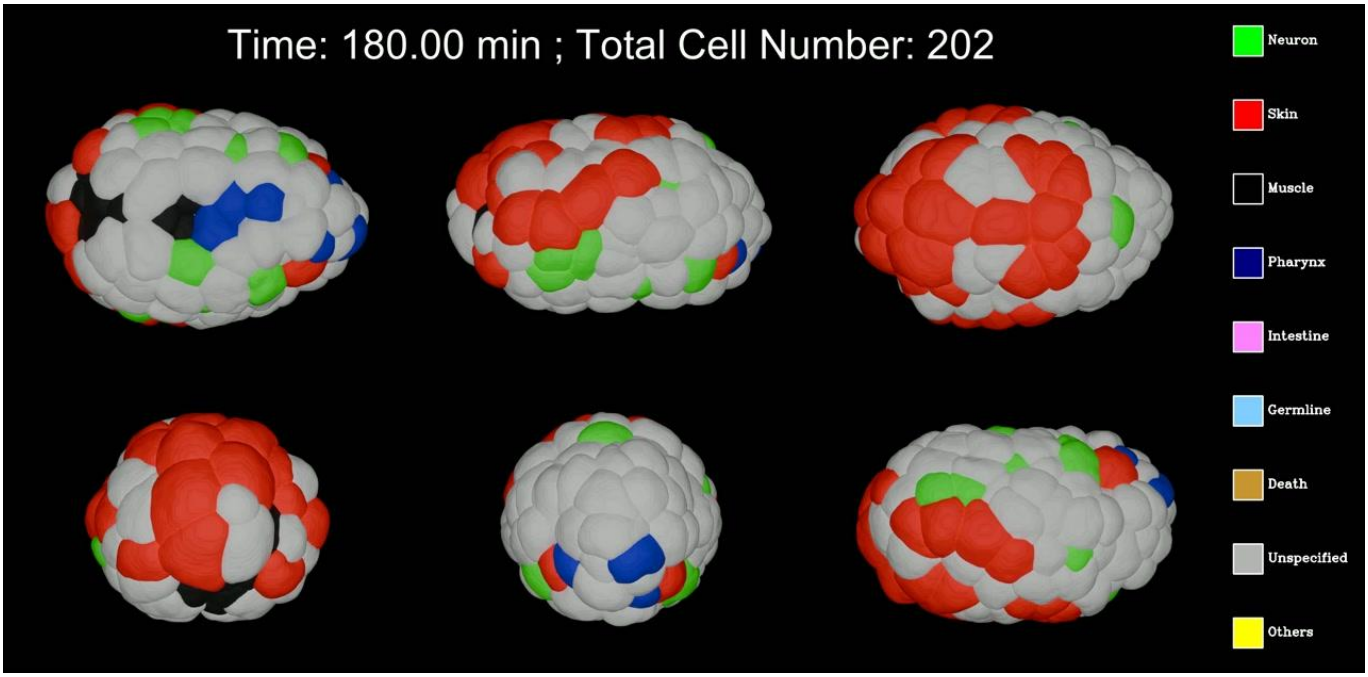
Figure S16. Surface unit cell adhesion on 30 sublineages.

Supplementary Movies



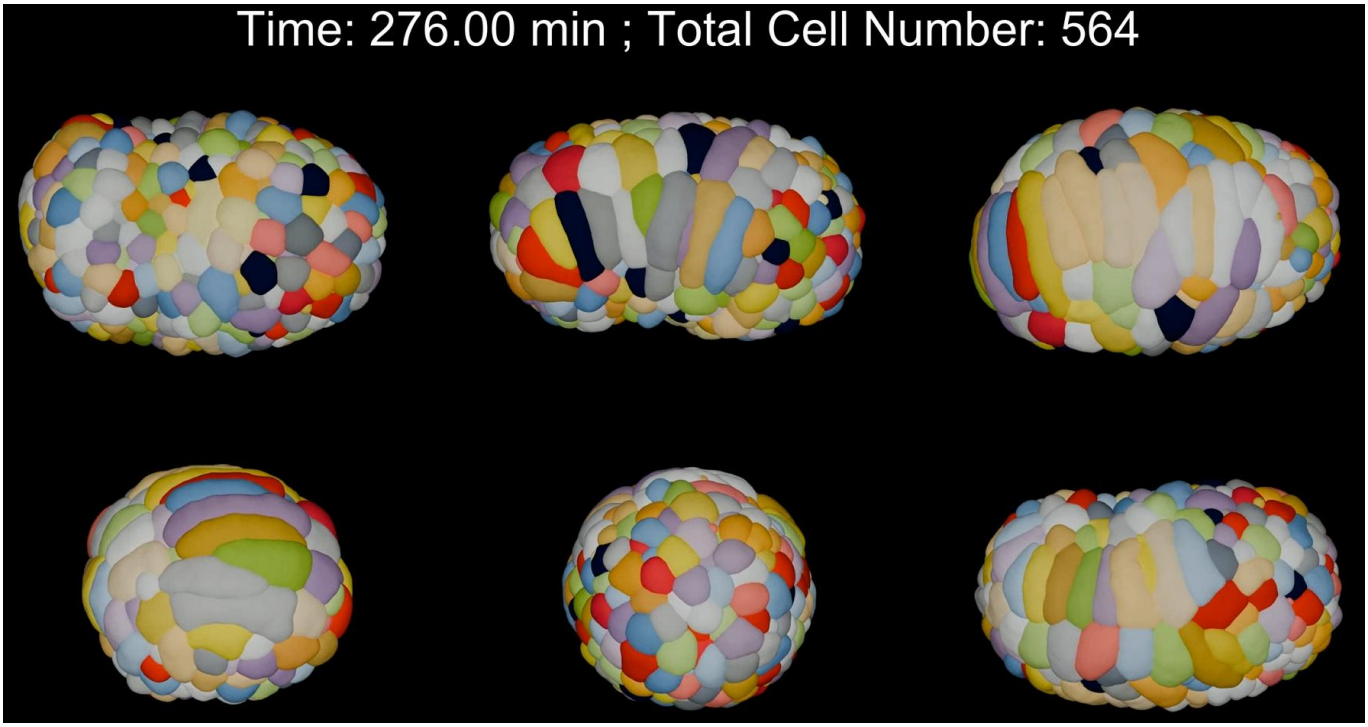
Movie S1. The cell-wise 4D timelapse development of *C. elegans* embryo 191108plc1p1.

This movie presents a four-dimensional (4D) timelapse visualization of *C. elegans* embryonic development. The recording begins at the 3-cell stage (t = 6.00 min) and documents development through to the 579-cell stage (t = 306.00 min). The process is captured across 201 consecutive time points at a 1.5-minute interval. To facilitate lineage tracing, each cell is computationally segmented and assigned a unique and consistent color throughout its history. The 3D model, rendered in Blender, is displayed simultaneously from six orthogonal viewpoints (dorsal, ventral, anterior, posterior, left, and right), providing a comprehensive and unambiguous spatial context for cellular dynamics.



Movie S2: Four-Dimensional Visualization of Cell Fate Specification in a *C. elegans* Embryo.

This movie presents a four-dimensional (4D) time-lapse visualization of embryonic development in *Caenorhabditis elegans* (embryo ID: 191108plc1p1), tracking cellular divisions and movements from the 3-cell to the 579-cell stage over 306 minutes. Cell lineages are computationally tracked, and each cell is assigned a distinct color based on its annotated terminal fate. This fate-based color map—which includes categories such as Neuron, Skin, Muscle, Pharynx, Intestine, Germline, Apoptotic, Unspecified, and Others—remains consistent throughout time, enabling direct observation of the spatiotemporal emergence and distribution of each lineage. Initial frames of the time-lapse showing the embryo at an early cleavage stage (3-cell stage at 6 minutes). To provide a comprehensive spatial context, the embryo is simultaneously rendered from six cardinal viewpoints (dorsal, ventral, anterior, posterior, left, and right) using Blender. This multi-perspective visualization provides a powerful resource for analyzing the intricate relationship between cell position, division history, and fate determination within the stereotyped developmental program of *C. elegans*. A later developmental stage where distinct cell fates, represented by different colors, have emerged and organized into clear patterns.



Movie S3: Dynamic Simulation of Tissue Morphogenesis using the *CTransformer*.

This video presents a 3D dynamic simulation of tissue morphogenesis driven by the *ctransformer* model, starting from an initial aggregate of 400 cells. The visualization is organized into a 2x3 grid, displaying cells as deformable, multi-colored polygons to clearly demarcate boundaries and track individual dynamics. On-screen metrics continuously monitor the elapsed simulation time and the expanding total cell population. As the simulation progresses, it captures the emergent structural evolution of the tissue caused by continuous cellular growth, mechanical pressure, and mitotic divisions. The multi-view format facilitates the comprehensive analysis of these complex spatial rearrangements, allowing for simultaneous observation across different model parameters or viewing angles. Overall, this visual rendering demonstrates the model's capability to generate realistic macroscopic tissue development based on fundamental, local cellular rules.

Supplementary Tables

Description of Table S1

Table S1 provides a comprehensive log of all embryo datasets utilized in this research. It is organized into two sections: "Uncompressed Embryos" and "Compressed Embryos," detailing both wild-type and mutant samples. The table catalogs critical metadata for each dataset, including imaging parameters like spatial and temporal resolution, maximum cell count, and segmentation dimensions.

Description of Table S2

This table presents the quantitative evaluation results for the *CTransformer* model's cell segmentation performance. The table lists individual embryo samples and reports their corresponding Intersection over Union (IoU) and Dice Score metrics. These scores serve as a direct measure of the model's accuracy in segmenting 3D cell objects compared to manually annotated ground truth.

Description of Table S3

This table details the segmentation performance of other models, which is used as a state-of-the-art benchmark for comparison. It provides the Intersection over Union (IoU) and Dice Score for cell segmentation on the same set of evaluation embryos as *CTransformer*. These data are crucial for contextualizing the performance advancements achieved by the *CTransformer* model.

Description of Table S4

The table offers a detailed, volume-wise evaluation of the *CTransformer* model's segmentation capabilities. It expands on the primary metrics by including Mean Surface Distance, the 95th percentile of Hausdorff Distance, Dice Score, Jaccard Index, and Cell Loss Rate for numerous embryo samples. This multi-faceted assessment provides a more granular understanding of the model's precision in capturing cell boundaries and its reliability in identifying all cells within a volume.

Description of Table S5

This table documents the performance of the *CTransformer* model in a time-lapse evaluation context. It contains a long list of embryo samples at numerous consecutive time points, reporting the Intersection over Union (IoU) and Dice Score for each. The purpose of this table is to demonstrate the model's temporal robustness and consistency throughout the dynamic process of embryonic development.

Description of Table S6

This table provides a quantitative evaluation of the StarDist3D segmentation method using the "3D Shape Evaluation" dataset. It lists the Intersection over Union (IoU), also known as the Jaccard Index, and the Dice Score for numerous embryo samples at various developmental time points. The scores exhibit considerable variability, with some samples showing high accuracy while others perform poorly.

Description of Table S7

This table details the segmentation performance of the CShaper++ algorithm on the ";3D Shape Evaluation" dataset. The metrics provided include the Intersection over Union (IoU) and the Dice Score for a comprehensive set of embryo images. The results demonstrate the capability of CShaper++ in cell segmentation, showing a range of accuracies across different developmental stages and samples.

Description of Table S8

This table presents a detailed analysis of the signal quality for both raw (real) and pseudo (generated) nuclei images from three wild-type embryo samples. It includes measurements of raw and pseudo signal power, noise power, and the calculated Signal-to-Noise Ratio (SNR) for each time point. This data provides a quantitative basis for comparing the images generated by the *m2nGAN* Module with the original fluorescence microscopy images.

Description of Table S9

This table documents the cell-wise evaluation of segmentation and tracking accuracy for the *CShaper++* method on the 170704plc1p1 embryo. It meticulously tracks the status of individual cells, identified by name, across numerous time points, marking each as 'Normal' (1) or lost (0). The table also specifies the developmental fate for many cells, such as intestine, muscle, or skin.

Description of Table S10

This table quantifies the reproducibility of cell surface adhesion measurements across three different HMR-1 embryo samples. For each identified cell at specific time points, it provides the average HMR-1 expression across replicates, the expression observed in the individual sample, and the resulting variation ratio. This analysis demonstrates the consistency and reliability of the *MolQuantifier* Module. The generally low variation ratios indicate that the *CTransformer* pipeline accurately and reproducibly measures cell-specific adhesion patterns, validating its use for quantitative biological analysis.

Description of Table S11

This table provides a quantitative comparison of cell lineage tracing performance using real nuclei fluorescence images versus pseudo-nuclei images generated by the m2nGAN module. The data includes accuracy, precision, and recall metrics for two wild-type embryo samples (WT_Sample4 and WT_Sample5) evaluated across numerous developmental time points.

Description of Table S12

This table presents the quantitative expression data for HMR-1/E-cadherin across three replicate *hmr-1::GFP* embryos. It details both the average and cumulative fluorescence intensity for various cell lineages, organized at two levels of resolution: 18 major sublineages and a more granular set of 36 sublineages.

Description of Table S13

This table contains the detailed measurements of adhesion asymmetry following cell division. The values represent the calculated difference in HMR-1/E-cadherin expression between the resulting daughter cells for numerous mother

cell lineages at various developmental time points. An additional summary provides the mean asymmetry value for each cell type across the dataset.

Description of Table S14

This table serves as a comprehensive reference list mapping cell names within the *C. elegans* embryonic lineage to their terminal developmental fates. It specifies the ultimate outcome for hundreds of cells, including differentiation into neurons, skin, muscle, intestine, or undergoing programmed cell death.

Description of Table S15

This table details the quantitative measurements of HMR-1/E-cadherin expression specifically at cell-cell contact interfaces. It lists the embryo-wise average fluorescence intensity for hundreds of unique contact pairs at precisely defined developmental time points.